



Lenguerrand, E., Marques, E., Artz, N., & Wylde, V. (2017, Sep 7).
ARENA-SHEAP: Statistical and health economic analysis plan. *Arena*.

Other version

License (if available):
Unspecified

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/pure/about/ebr-terms>

ARENA-SHEAP

Statistical and health economic analysis plan

This document describes the analytical strategy for the ARENA trial study.

It has been developed by

Dr Erik Lenguerrand (Medical Statistician)

Dr Elsa Marques (Health Economist)

Dr Neil Artz (Clinical Lead)

Dr Vikki Wylde (Chief Investigator)

on behalf of the ARENA research team.

Table of Contents

I.	Data cleaning and quality control checking	4
II.	Deriving the primary and secondary outcomes	7
a.	Lower Extremity Functional Scale (LEFS).....	7
b.	Knee injury and Osteoarthritis Outcome Score (KOOS)	8
c.	HADS.....	9
d.	Self-Administered Patient Satisfaction Scale for Primary Hip and Knee Arthroplasty	10
e.	EQ-5D	10
f.	Resource use categories	11
III.	Effectiveness analyses	12
a.	Participant flow	12
b.	Comparison of baseline characteristics	12
c.	Primary analysis and related analyses.....	12
i.	Intention-to-treat analysis and dealing with missing primary outcome data	13
i.	. Primary analysis (LEFS at 12 months post-operative)	14
ii.	Sensitivity analyses	16
iii.	Additional analyses	17
iv.	Secondary analyses	19
IV.	Cost-effectiveness analyses.....	23
a.	Objective	23
b.	Measurement of resources	23
c.	Defining missing resource use categories	24
d.	Valuation of resources	25
e.	Estimation of costs and QALYs.....	25
f.	Dealing with missing data.....	26
g.	Analysis.....	27
h.	Sensitivity analysis.....	29
i.	Presentation of results	29
V.	Other analyses.....	30
a.	Serious adverse events	30
b.	Evaluation of the intervention.....	30
c.	Further exploratory investigation.....	31

I. Data cleaning and quality control checking

An ongoing data quality checking process will be implemented throughout the course of the trial. For the first 25 recruited patients, a double-data entry exercise will be conducted for all the information already imputed in the database. During the trial, a random sample of 5 patients will be selected at regular intervals and all their completed questionnaires will be double-entered. Fields with the highest errors frequencies will be identified and appropriate actions implemented including re-training staff.

Continuous double data entry of pre-operative and 12 months post-operative LEFS and EQ-5D will be conducted for each patient once their 12 months questionnaire has been completed.

During the trial, research nurses, the trial manager and trial administrative support staff reviews responses to the questionnaires. When questions are incomplete or the patients fail to return the questionnaires, research staff follows patients up with a telephone call. During this telephone call, research staff queries missing responses to the questionnaire. For non-returners, research staff delivers a shorter version of the full follow-up form which includes only the primary clinical outcome tool (LEFS) and the primary economic result tools (EQ-5D and resource use questionnaire).

Towards the end of the trial (months 23-27), a thorough data cleaning exercise will be conducted. The fields of interest for the analyses presented in section II, III and IV will be investigated to identify 1.incomplete items, 2.implausible values and if relevant 3.inconsistencies across measurement points/questionnaire.

For categorical variables, if participants put a mark outside the box associated with one item (category), the closest box will be assigned. If two boxes are marked, that which indicates the more severe problem will be chosen.

Categorical/binary variables will be tabulated to identify implausible coding/categories and missing items. The database should tag all the fields not completed by the participants, allowing the identification of fields accidentally missed during the data inputting process. For missing items and implausible code, the original data will be checked on the completed questionnaires/case report forms and any data entry errors corrected. In case of incertitude, the Chief Investigator/trial manager and study statistician/ health economists will take a collegial decision to amend the field appropriately. For remaining implausible values, a

variable that indicates ‘possible implausible value of X’ (where X is the name of the variable that has a possible implausible value) will be derived and the implausible value in X will be discarded.

Continuous variables will be investigated with histograms, box and/or normal plots. Values that appear outside of the main distribution in the majority of participants (i.e. outliers) on normal plots and histograms, as well as outside the “whiskers” of the box plot will be considered implausible. For implausible values, the original data will be checked on the completed questionnaires/case report forms and any data entry errors corrected. In case of incertitude, the Chief Investigator/trial manager and study statistician will take a collegial decision to amend the field appropriately. For remaining implausible values, a variable that indicates ‘possible implausible value of X’ (where X is the name of the variable that has a possible implausible value) will be derived. In the primary and secondary effectiveness analyses we will complete analyses with all participants included (including where they have a possible implausible value) and again with participants excluded for analyses with a given outcome if their value for that outcome has been marked as possibly implausible. If removal of participants with possible implausible values results in a change of a magnitude that would affect the interpretation/conclusion for that outcome then both sets of results will be reported; otherwise only the results with all included irrespective of ‘implausible value’ status will be reported.

Composite scores derived by adding/combining discrete questionnaire items (see following section) will be investigated in a similar manner as continuous variables. We do not expect implausible total scores as the logical rules implemented in the data entry forms should prevent such a situation. However, if this were to be encountered, scores outside the expected range will be considered as implausible and discarded.

The cost-effectiveness analyses will follow the effectiveness analyses in terms of participants excluded and will use the outcome values derived in the effectiveness analyses to compute its results. For resource use categories, we will investigate both the range of values reported and the patterns of missing data per time point. Implausible very high or negative values will be first checked against the primary data collection source (questionnaire or case report form); and then checked against patterns of the same or related resource use categories and sociodemographic variables. If values are still deemed implausible on final check, they will

be discarded. Discarded values would therefore be considered missing values and completed using multiple imputation methods (see section IV.f.).

II. Deriving the primary and secondary outcomes

a. Lower Extremity Functional Scale (LEFS)

The primary outcome is the LEFS, a composite score derived by summing together 20 Likert-scale type items. Those items relate to four different groups of activities: Hardest Activities, Moderately Difficult Activities, Moderately Easy Activities and Easy Activities (see table below extracted from Stratford PW, Hart DL, Binkley JM, Kennedy DM, Alcock GK, Hanna SE: Interpreting Lower Extremity Functional Status Scores. *Physiother Can* 2005, 57(2):9.) Each item ranges from 0 (extremely difficult) to 4 (no difficulty) and the total score ranges from 0 (high disability) to 80 (no disability). Up to 4 missing item responses are permitted to derive the total score.

The guidelines for handling missing items are as follows:

1. A minimum of 16 items must be answered to use the LEFS
2. No more than two missing items can come from any one of the groups reported in Table 2.
3. To impute an item score, average the item values provided by the patient for the two closest items identified in the item hierarchy in Table 2.
4. If the most difficult item response (running on uneven ground) is missing, assign the score of the closest answered item.
5. If the easiest item response (walking between rooms) is missing, assign the score for the closest answered item.

To illustrate these steps, suppose that a patient left the item "getting in or out of the bath" blank. One would impute a value for this item by averaging the responses for "getting in or out of a car" and "performing light activities around the home." Thus, if the respective scores were 3 and 2, the imputed item score would be 2.5.

Stratford et al Interpreting Lower Extremity Functional Status Scores

Table 2 LEFS Item Difficulty Hierarchy

<i>Item Grouping</i>	<i>Response Options and LEFS Difficulty Values (LEFS Item Response-LEFS Total Score)</i>
Group 1: Hardest Activities	
Running on uneven ground	(0-28; 1-45; 2-59; 3-67; 4-76)
Making sharp turns while running fast	(0-28; 1-47; 2-59; 3-65; 4-75)
Running on even ground	(0-29; 1-44; 2-56; 3-65; 4-74)
Hopping	(0-26; 1-43; 2-54; 3-63; 4-73)
Walking a mile	(0-21; 1-38; 2-46; 3-60; 4-71)
Group 2: Moderately Difficult Activities	
Your usual hobbies, recreational or sporting activities	(0-20; 1-32; 2-45; 3-59; 4-71)
Squatting	(0-21; 1-35; 2-46; 3-56; 4-68)
Standing for 1 hour	(0-19; 1-33; 2-45; 3-58; 4-69)
Performing heavy activities around your home	(0-17; 1-32; 2-45; 3-57; 4-71)
Going up or down 10 stairs	(0-16; 1-38; 2-41; 3-54; 4-69)
Group 3: Moderately Easy Activities	
Any of your usual work, housework, or school activities	(0-13; 1-25; 2-39; 3-54; 4-70)
Walking 2 blocks	(0-15; 1-27; 2-37; 3-49; 4-66)
Lifting an object, such as a bag of groceries from the floor	(0-11; 1-24; 2-33; 3-47; 4-61)
Getting into or out of a car	(0-7; 1-17; 2-29; 3-45; 4-63)
Getting into or out of the bath	(0-11; 1-18; 2-30; 3-42; 4-59)
Group 4: Easy Activities	
Performing light activities around your home	(0-6; 1-17; 2-27; 3-43; 4-63)
Sitting for 1 hour	(0-12; 1-26; 2-33; 3-45; 4-53)
Rolling over in bed	(0-14; 1-20; 2-30; 3-42; 4-56)
Putting on your shoes or socks	(0-8; 1-18; 2-28; 3-41; 4-57)
Walking between rooms	(0-4; 1-15; 2-22; 3-36; 4-57)

LEFS = Lower Extremity Functional Scale.

This is a score in whole numbers and will be treated as a continuous variable.

b. Knee injury and Osteoarthritis Outcome Score (KOOS)

The KOOS is a secondary outcome. It consists of 5 subscales; pain, other symptoms, function in daily living (ADL), function in sport and recreation (Sport/Rec) and knee related quality of life (QOL). Standardized answer options are given (5 Likert boxes) and each item is assigned a score from 0 (none) to 4 (extreme). A normalized score (100 indicating no symptoms and 0 indicating extreme symptoms) is calculated for each subscale:

1. PAIN $100 - \frac{\text{Mean Score (P1-P9)} \times 100}{4} = \text{KOOS Pain}$
2. SYMPTOMS $100 - \frac{\text{Mean Score (S1-S7)} \times 100}{4} = \text{KOOS Symptoms}$
3. ADL $100 - \frac{\text{Mean Score (A1-A17)} \times 100}{4} = \text{KOOS ADL}$
4. SPORT/REC $100 - \frac{\text{Mean Score (SP1-SP5)} \times 100}{4} = \text{KOOS Sport/Rec}$
5. QOL $100 - \frac{\text{Mean Score (Q1-Q4)} \times 100}{4} = \text{KOOS QOL}$

As long as at least 50% of the subscale items are answered for each subscale, a mean score can be calculated. If more than 50% of the subscale items are omitted, the response is considered invalid and no subscale score should be calculated. For the subscale Pain, this means that 5 items must be answered; for Symptoms, 4 items; for ADL, 9 items; for Sport/Rec, 3 items; and for QOL, 2 items must be answered in order to calculate a subscale score. Subscale scores are independent and can be reported for any number of the individual subscales, i.e. if a particular subscale is not considered valid (for example, the subscale Sport/Rec 2 weeks after total knee replacement), the results from the other subscales can be reported at this time-point.

A total score has not been validated and is not recommended. For the purpose of an RCT, KOOS subscale scores can be aggregated and averaged as a primary outcome. For the purpose of this trial, we will investigate the averaged KOOS and the five individual KOOS subscale scores to enable clinical interpretation. If any subscale scores are missing, the averaged score will not be derived.

These are scores in whole numbers and will be treated as continuous variables.

c. HADS

The HADS is a secondary outcome. It is comprised of two subscales: Depression and Anxiety. Each sub-scale is derived by summing together 7 Likert-scale type items. Items are rated on a 4-point Likert-type scale ranging from 0 to 3, with higher scores representing greater symptom severity. Each subscale has a score ranging from 0 to 21.

It is recommended that the score for a single missing item from a subscale is inferred by using the mean of the remaining six items. If more than one item is missing, then the subscale should be judged as invalid. We will therefore consider the subscale as missing if more than one item is missing.

These are scores in whole numbers and will be treated as continuous variables.

d. Self-Administered Patient Satisfaction Scale for Primary Hip and Knee Arthroplasty

The Patient Satisfaction Scale is a secondary outcome. It is comprised of four items. Each item is scored on a 4-point Likert scale with response categories consisting of very satisfied (100 points), somewhat satisfied (75 points), somewhat dissatisfied (50 points), and very dissatisfied (25 points). The scale score is the unweighted mean of the scores from the individual items, ranging from 25 to 100 per item (with 100 being most satisfied).

There is no published recommendation on how to handle missing item information. We will therefore consider the satisfaction scale as missing if any item is missing.

This is a mean score with continuous values between 25 and 100 and will be treated as a continuous variable.

An unvalidated single-item satisfaction outcome is also considered to assess the satisfaction with the physiotherapy received by both groups of participants. Based on a 5-item Likert scale ranging from very satisfied (1) to very dissatisfied (5), this variable describes satisfaction with physiotherapy treatment. This is a categorical variable and will be treated as an ordinal variable.

e. EQ-5D

The EQ-5D-5L response profiles will be combined using weights from the EuroQol's published UK societal utility tariffs to produce a composite quality of life score at every time point (2-weeks, and 3-, 6-, and 12-months post-operative) for each patient. These quality of life, or "utility" scores, are used to derive quality adjusted life years, or QALYs.

The EQ-5D-5L QoL scores will be treated as continuous variables and are limited at a maximum of 1 (corresponding to perfect health), where 0 corresponds to death. Negative values are permitted (worse than death).

In economic evaluation, outcomes, such as the QALY, should be jointly imputed with costs. A QALY in this study will be computed based on the 5 domains of the EQ-5D-5L administered over 4 time points, for a total of up to 20 variables. If any of the 5 items within the EQ-5D-5L are missing, the utility score will be recorded as missing in a separate indicator variable, per time point, per patient. In section V.f. “Dealing with missing data” we explore strategies to deal with EQ-5D-5L missing data.

f. Resource use categories

For resource use categories, the range values reported per time point will be investigated. We will look at patterns of missing data per time point. Implausible high values will be rounded to the maximum possible value within the period of resource use data collection; or discarded if the pattern of missing data for a particular resource category is unclear. Discarded values would therefore be considered missing values and completed using multiple imputation methods.

Research data entry staff noted patterns of missing or poorly completed data during data entry of resource use questions which prompted telephone checks. The format of resource use questions was amended March 2017 to improve self-completion. This review also prompted a question on admission to nursing or rehabilitation homes to be separate from other secondary care admissions. A field in the database will be created to distinguish between patients completing versions before and after this format change. We will check patterns of completion for the questions before and after the format changes. If more data is reported after the change in a substantial number of cases, we will decide between:

- excluding this category for economic analysis,
- imputing the data in the initial version if the imputation model can efficiently estimate this,
- consider all data as is, if not likely to bias or impact results.

Methods to deal with missing data, including the multiple imputation model are detailed in section IV.f. “Dealing with missing data”. Imputation models for resource use and outcome variables can be very complex and adding layers of complexity may not be efficient.

III. Effectiveness analyses

a. Participant flow

A CONSORT diagram will summarise participant flow through the trial, documenting invitation and recruitment, receipt of intervention or usual care as allocated, and collection of data (pre-operatively, during the 12 weeks following surgery and at 3 months, 6 months, and 12 months post-operatively). We will define the recruitment rate as the percentage of eligible patients who are recruited and successfully randomised.

The number and percentage of participants who withdraw, with and without primary outcome data, at each of the follow-up stages will be reported for each trial group (intervention or usual care) and compared using chi-squared or Fisher-exact tests.

b. Comparison of baseline characteristics

We will compare relevant summary statistics of baseline characteristics between participants allocated to the intervention group and usual care group in order to determine whether any potentially influential imbalance has occurred (by chance) between these two groups. We will focus on demographic, surgical and stratification variables, as well as outcome measures at baseline.

The comparisons between the two groups will be made by summarising variables in each group (intervention versus usual care groups). For all continuous and score variables we will check distributions using histograms and normal plots to examine how close to normality these are before deciding which summary statistics to present. Continuous variables with approximately normal distributions will be presented as means and standard deviations (SD). Continuous variables that we anticipate will not have an approximate normal distribution will be presented as medians and interquartile ranges (IQR). Binary/categorical variables will be presented as number (n=) and percentage (%).

We will not compare baseline characteristics between the two groups with a statistical test (p-value) as any low values simply represent a type-1 error under the assumption that we have adequately randomly allocated participants.

c. Primary analysis and related analyses

The primary outcome is the LEFS score measured at 12-months post-operative. This study has been designed and powered to detect a difference equivalent to a minimal clinically

important difference of 9 scale points between the two groups with a power of 80%, a two-sided 5% significance level and accounting for up to 35% missing data.

i. Intention-to-treat analysis and dealing with missing primary outcome data

The primary analysis will be conducted on an intention-to-treat (ITT) basis with due emphasis placed on the confidence interval for the between-group comparison when comparing the primary outcome for the two groups (intervention versus usual care groups).

ITT requires all participants in a clinical trial to be included in the main analyses in the groups to which they were randomised. Compliance with the ITT principle requires a strategy to handle loss to follow-up or missing data on primary outcome at follow-up amongst those who have been randomised.

The primary analysis will be based on all randomised participants with available LEFS outcomes scores at at least one post-operative assessment point. In other words, the analysis will include patients with LEFS scores available at every post-operative follow-up assessments, patients with scores available only at the 18 months assessment (primary outcome) and those with missing primary outcome but at least one other post-operative assessment. The data will be analysed with multilevel linear regression, a likelihood-based method which will provides an unbiased estimate of the intervention effectiveness (at 18 months) by including all available post-randomisation data for participants with missing primary outcome data.¹ This approach relies on the data being missing at random (MAR) and does not include the randomised patients with no post-operative LEFS measures at any post-operative time point.

To include all randomised participants in respect with the ITT principle we will also use a multiple imputation strategy (chained equations). The imputation process will be stratified by trial group. The number of imputation sets will be determined by the number of patients with missing outcomes. For example, if 14% of patients have missing LEFS at 12-months, 14 imputations will be performed. The variables used on the imputation process will be the LEFS scores measured at 3-, 6- and 12-months post-operative, the stratification variables (the pre-operative LEFS and a factor identifying the recruitment centre), the imbalanced baseline characteristics between trial groups used in the adjusted model, and relevant ancillary variables. The ancillary variables will consist of the baseline comorbidities, surgical and socio-demographics characteristics which differ between participants with complete and

missing primary outcome. (The selection process of these ancillary variables will also identify the pattern of randomness (MAR or Missing Completely At Random) of the studied data). The Rubin's rules will be used to combine the resulting multiple imputed estimates. This strategy relies on the MAR hypothesis.

The following strategies will also be used to assess the sensitivity of the above results to departures from MAR as described by White et al.^{1, 2, 3}. The mean of the unobserved responses for LEFS at 12 months will be considered to be 10% worse than the mean of the observed responses (irrespective of randomised group). We will also assumed that the data will be informatively missing only in the intervention arm, with the mean of the unobserved responses being as much as 10% more and then 10% less than the mean of the observed responses, and comparable between unobserved and observed responses in the control group. Patients who withdraw will not be considered in the analyses.

i. . Primary analysis (LEFS at 12 months post-operative)

The primary outcome, the LEFS at 12 months post-operative, is a score treated as a continuous variable. We will first use a normal plot and a histogram to assess the normality of this measure. If the outcome is approximately normally distributed, the raw score will be used with no transformation. If important departure from the normal distribution is observed, we will explore transforming the score to improve the normality of the residuals in the regression model. The decision to transform the score or not, and if so which transformation to use, will be decided by considering: (1) the distribution of the variable, (2) the distribution of residuals from regression models, (3) the ease of interpreting results following any given transformation compared with no transformation and (4) whether main results/conclusions are influenced by the transformation or not. If the overall conclusion is not altered by whether the variable is transformed or not, we would use the untransformed (easier to interpret) outcome.

We will use a linear mixed regression to model the difference in means LEFS scores between participants allocated to the intervention group and those allocated to the usual care group (reference group), whilst taking account of clustering induced by the repeated measures within patients. We will model the LEFS scores measured at 3-, 6- and 12-months. As described in

1. Including all individuals is not enough: lessons for intention-to-treat analysis. White IR, Carpenter J, Horton NJ. Clin Trials. 2012 Aug;9(4):396-407. doi: 10.1177/1740774512450098.

2.Strategy for intention to treat analysis in randomised trials with missing outcome data. White IR, Horton NJ, Carpenter J, Pocock SJ. BMJ. 2011 Feb 7;342:d40. doi: 10.1136/bmj.d40.

3. Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial. White IR, Kalaitzaki E, Thompson SG. Stat Med. 2011 Nov 30;30(27):3192-207. doi: 10.1002/sim.4360. .

section III.C.i, this model allows the production of estimates which are valid under the MAR hypothesis and therefore avoids imputing the outcomes for those with missing information whilst complying with the ITT principle. The model will be adjusted for the stratification variable used in the randomisation, i.e. the pre-operative LEFS score (model as a continuous variable) and a factor identifying the recruitment centre (Southmead vs Emerson Green).

The main model will be

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{1ij} X_{2ij} + \beta_5 X_{1ij} X_{3ij} + \beta_6 X_{6ij} + \beta_7 X_{7ij} + u_{0j} + e_{0ij}$$

Where

Y_{ij} is the LEFS outcome at time $i=1$ (3 months post-operative), 2 (6 months post-operative) or 3 (12 months post-operative) for participant $j=1 \dots 256$.

β_0 is the intercept, i.e. the outcome at 12 months amongst those in usual care group, operated on at Southmead Hospital, with the lowest level of pre-operative LEFS score.

β_1 is the intervention effect, i.e. the mean difference in LEFS outcome at 12 months comparing participants in the intervention group to those in the usual care group (reference), having adjusted for recruitment centre, pre-operative LEFS score and taking into account the non-independence amongst measurements from the same participant (repeated measurements). X_{1ij} is the factor coded 1 if the participant received the intervention, 0 if the participant had only the usual care (reference). In the primary analysis, the focus will be only on this estimate, its 95%CI interval and p-value.

β_2 and β_3 are respectively the effect associated with 3 months X_{2ij} and 6 months X_{3ij} assessment points covariables among participants in the usual care group, i.e. the mean difference in LEFS outcome between respectively the 3 months and 12 months LEFS scores and then between the 6 months and 12 months scores among participants in the usual care group. The 12 months assessment point is used as the reference and not model per se in the equation.

β_4 and β_5 are the interaction coefficient for the interaction of the intervention indicator with, respectively, the 3 months or 6 months assessment points covariables. Those parameters will be used to identify the intervention effect at 3 months and 6 months.

β_6 and β_7 are the effects associated with the stratifying covariables X_{6ij} and X_{7ij} , respectively the LEFS pre-operative score and the recruitment centre (Southmead Hospital as a reference).

u_{0j} is the residual of the outcome for participant j .

e_{0ij} is the residual of the outcome for measurement i of participant j .

For the effect associated with β_1 , statistical significance will be indicated by a two-sided p-value of ≤ 0.05 .

ii. Sensitivity analyses

Different sensitivity analyses will be conducted.

The first analysis will investigate the impact of clustering of outcome by surgeon that performed the operation on the intervention effect at 12 months, i.e. the amount of outcome variability explained at surgeon level rather than individual level. A third level (patients nested within surgeon) will be added to the equation presented in the previous section. If the variance of the random effect associated with surgeon level is significant, this level will be kept for the following sensitivity analyses; otherwise the two-level model presented in the previous section will be used.

The following analysis will adjust the intervention effect for any baseline imbalanced participant characteristics. Given the published differences in outcomes after TKR for men and women, exploratory analysis will be undertaken to investigate the impact of gender, i.e. the analysis will be first adjusted for gender and the interaction between the intervention effect and gender will then be explored. The trial is not powered for such adjustments and it will only inform us on their potential impact on the intervention effect.

The main analysis as well as the above sensitivity analysis will be replicated using the different scenarios to handle missing LEFS outcome presented in section III.c.i..

Finally the main and adjusted sensitivity analyses (without imputation) will be performed using a per-protocol (PP) approach. Participants randomised to the intervention group and who attended less than 4 intervention sessions (<4) will be considered as non-compliant. They will be modelled in the group they were randomised to for the ITT-analyses but will be excluded from the PP models. We do not anticipate that patients allocated to the usual care group will attend intervention session but if this were the case (at least four sessions), they would still be modelled in the control group for the ITT-analyses but would be excluded from

the PP models. Further exploratory analyses will be conducted to investigate how change in the number of attended session(s) used to define the threshold for intervention adherence will affect the results.

For the adjusted sensitivity analyses, the adjustment covariates with missing information will be handled using the missing indicator strategy⁴. In addition to the covariate of interest, an indicator variable (indicating whether a baseline covariate is missing or observed) will be introduced in the equation presented above together with imputing the missing covariate with a ‘temporary’ value derived from the observed baseline covariate values (usually the mean). The adjustment covariates with missing information will also be imputed chained equations in the multiple imputation model described in section III.c.i.

iii. Additional analyses

If no transformation of the LEFS score at 12 months post-operative provides an approximately normally distributed outcome and regression residuals, the main analysis presented above will be reported, i.e. the raw LEFS outcome will be modelled with the linear mixed regression described above. The resulting intervention effect will be contrasted with the following results:

- The LEFS median score at 12 months and interquartile-range will be presented by trial group and compared with a Wilcoxon-Mann-Whitney test.

- The LEFS outcome will be transformed into an ordinal dependent variable and modelled with a multilevel ordered logistic regression or a multilevel generalized ordered logit model. At the time of writing this SHEAP there is no clear guidance on how to categorise the LEFS outcome. We propose to use the following categories:

[0-20], [21-40], [41-60], >60. That categorisation will be revisited if new guidance are published before the end of this trial. The final choice of categorisation will be approved by the Trial Steering Committee who will consult an external statistician to validate the decision.

The model will provide a sets of odd-ratios associated with the intervention effect:

- comparing the odd of having a LEFS score of {[0-20]} vs the odds of having a LEFS score of {[21-40], [41-60], >60} for those who received the intervention vs those who did not

- the odd of having a LEFS score of {[0-20] , [21-40]} vs the odds of having a LEFS score of

4. Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis. Groenwold RH, White IR, Donders AR, Carpenter JR, Altman DG, Moons KG. CMAJ. 2012 Aug 7;184(11):1265-9. doi: 10.1503/cmaj.110977

{ [41-60], >60} for those who received the intervention vs those who did not

- the odd of having a LEFS score of { [0-20] , [21-40],[41-60] } vs the odds of having a LEFS score of { >60} for those who received the intervention vs those who did not

We will also investigate the influence of additional physiotherapy treatment received by participants. The main analysis will be adjusted for use of additional physiotherapy.

Similarly, although each physiotherapy class will be delivered by two physiotherapists, the physiotherapists delivering the classes are likely to vary. In an exploratory analysis, we will investigate if the mean/median LEFS score at 12 months post-operative differed according to the lead physiotherapist in charge of the attended intervention sessions. As participants could be supervised by several physiotherapists through the course of their 6 sessions, they will be related to the lead physiotherapist who will have delivered most of their sessions. If any difference in LEFS mean/median between physiotherapists is identified, these evidences will be used to discuss the intervention effect assessed in the main analysis.

iv. Secondary analyses

LEFS 3 months and 6 months post-operative

Using the same models presented below, we will estimate the intervention effect on LEFS at 3 months and 6 months, respectively $\beta_1 + \beta_4$ and $\beta_1 + \beta_5$. We will also determine if:

- the intervention effect at 3 months differs from the intervention effect at 12 months: β_4 ;
- the intervention effect at 6 months differs from the intervention effect at 12 months: β_5 ;
- the intervention effect at 3 months differs from the intervention effect at 6 months: $\beta_5 - \beta_4$

We will focus on the 95% CI and p-values associated with coefficients or linear combinations.

KOOS

The KOOS subscales will be collected pre-operatively, and at 3 months, 6 months and 12 months post-operatively. They are scores and will be treated as continuous outcomes.

Similar to the LEFS scores, those outcomes will be modelled with a linear mixed model. The effect of the intervention on the 5 KOOS sub-scales and on the average total KOOS score will be investigated in separate analysis sets, i.e. 6 different sets of analysis in total.

The strategy to handle missing outcomes will be similar except that each KOOS subscale and the average total KOOS score will be imputed with separate process. The multiple imputation strategy will use the post-operative KOOS scores rather than the post-operative LEFS scores to impute missing scores and will also use the pre-operative KOOS score in addition to the pre-operative LEFS score and relevant baseline comorbidities and socio-demographics characteristics.

A similar modelling strategy as the one presented above for the LEFS score will be used:

- First, an ITT-regression similar to the main analysis for the primary outcome but further adjusted for $\beta_8 X_{8ij}$, the pre-operative KOOS sub-scale score/ average total KOOS score;
- the ITT-adjusted model;
- the different sensitivity analyses for handling missing outcome
- the main analysis and adjusted analyses will be repeated on a PP-basis.

The raw scores will be modelled but if required appropriate transformation will be considered. If no suitable transformation is found, the scores will be categorised using appropriate thresholds according to published and accepted evidences and in the absence of such evidences we will consider the following classification: [0-25], [26-50], [51-75], >75.

The final choice of categorisation will be approved by the Trial Steering Committee who will consult an external statistician to validate the decision.

We will first focus on the intervention effect on KOOS at 12 months, then at 3 and 6 months and then consider the difference in intervention effect between each measurement point as described previously for the LEFS scores.

HADS

The HADS subscales will be collected pre-operatively, and at 3 months, 6 months and 12 months post-operatively. They are scores and will be treated as continuous outcomes. A similar modelling strategy as for LEFS and KOOS scores will be considered using linear mixed regressions. Two sets of analyses will be conducted; one for each subscale of the HADS.

The strategy to handle missing outcomes will be similar except that each HADS subscale will be imputed with separate process. The multiple imputation strategy will use the post-operative HADS scores rather than the post-operative LEFS scores to impute missing scores and will also use the pre-operative HADS subscale score in addition to the pre-operative LEFS score and relevant baseline comorbidities and socio-demographics characteristics.

The analyses will be further adjusted by $\beta_8 X_{8ij}$, the pre-operative HADS subscale score (the pre-operative KOOS scores will not be considered in the analyses of HADS subscale scores).

The raw scores will be modelled but if required appropriate transformation will be considered. If no suitable transformation is found, the scores will be categorised using the following published thresholds: [0-7], [8-10], [11-21]. If the sample size in category [11-21] is not large enough, categories [8-10] and [11-21] will be collapsed.

We will first focus on the intervention effect on HADS at 12 months, then at 3 and 6 months and then consider the difference in intervention effect between each measurement point as described previously for the LEFS scores.

Satisfaction with knee replacement and physiotherapy treatment

The satisfaction with surgery score will be collected at 3-months, 6-months and 12-months post-operatively and will be treated as a continuous outcome. A similar modelling strategy as for LEFS scores will be considered using linear mixed regressions.

The strategy to handle missing outcomes will be similar as for LEFS scores. The multiple imputation strategy will use the post-operative satisfaction with surgery score rather than the post-operative LEFS scores to impute missing scores and will use the pre-operative LEFS score and relevant baseline comorbidities and socio-demographics characteristics.

The analyses will not be further adjusted by a pre-operative score as this is irrelevant for this outcome. The raw scores will be modelled but if required appropriate transformation will be considered. If no suitable transformation is found, the scores will be categorised using the following published thresholds: [25-50], [51-75], >75.

The satisfaction score with the physiotherapy received is an ordinal variable also collected in each post-operative questionnaire. It will be considered with a multilevel ordered logistic regression or a multilevel generalized ordered logit model (depending how the hypothesis of proportional odds-ratio hypothesis holds true). A similar modelling strategy as for the above outcomes will be considered.

The multilevel ordered logistic model will be

$$\begin{aligned} \text{logit}\{\Pr(Y_{ij} > s | X_{ij}, u_{0j})\} \\ = \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{1ij} X_{2ij} + \beta_5 X_{1ij} X_{3ij} + \beta_6 X_{6ij} + \beta_7 X_{7ij} + u_{0j} - K_s \end{aligned}$$

Where Y_{ij} is the ordinal outcome with S ordinal categories denoted s (S=1,...,5). The observed ordinary satisfaction variable Y_{ij} is generated from the latent continuous response y_{ij}^* via a threshold model:

$$Y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* \leq K_1 \\ 2 & \text{if } K_1 < y_{ij}^* \leq K_2 \\ 3 & \text{if } K_2 < y_{ij}^* \leq K_3 \\ 4 & \text{if } K_3 < y_{ij}^* \leq K_4 \\ 5 & \text{if } K_4 < y_{ij}^* \end{cases}$$

Where K_s are the category-specific thresholds (i.e. the category of the variable);

Where the u_{0j} is a patient-specific random intercept, the overall level or intercept of the cumulative logits which varies over patients j ;

And where the β and X have the same meaning as previously.

We will first focus on the intervention effect on satisfaction with surgery or satisfaction with physiotherapy scores at 12 months, then at 3 and 6 months and then consider the difference in intervention effect between each measurement point as described previously for the LEFS scores.

IV. Cost-effectiveness analyses

a. Objective

To compare the physiotherapy intervention versus standard care in relation to costs and outcomes at 12 months after knee replacement surgery.

b. Measurement of resources

Resources used in relation to the delivery of the intervention will be recorded in study report forms. The physiotherapists delivering the sessions are asked to record time spent preparing and delivering the session, and the number of patients attending the session. The physiotherapist will also collect travel expenses for the patients attending sessions in a travel expenses form. These forms will collect information that will be used to value patients' private travel expenses associated with the intervention.

Resource use in relation to the knee replacement surgery, knee pain or knee function will be measured from randomisation until last follow-up. Resource use in relation to their knee surgery and function will be collected from patient self-completed questionnaires, completed at 3, 6 and 12 months after the knee replacement operation. Resources used within 2 weeks of the operation are not collected because randomisation occurred at 2 weeks; hence, no difference in resource use in relation to the intervention is expected between groups

Patients are supplied with resource use logs at 2-weeks post operation (not analysed), to help them complete the subsequent questionnaires. Patient resource use is collected on the following categories:

NHS resources

- a) inpatient and day case secondary care visits;
- b) A&E and outpatient visits, including hospital physiotherapy provided;
- c) community based NHS visits, such as GP and nurse visits; community physiotherapy and occupational therapy visits;
- d) medication use;
- e) occupational therapy;

Personal Social Services

- a) food at home services;
- b) home care help services;

- c) social worker contacts;
- d) equipment provided to patients and changes made to their homes.*

* Note that some of equipment is provided by NHS occupational therapists at hospital discharge but paid by social services. On other occasions, these costs may be borne by the NHS. We will liaise with the finance department for both centres to confirm the nature of this resource use category for our patient group (NHS or PSS).

Private expenses

- a) patient travel to the intervention;
- b) private healthcare or therapy;
- c) privately bought equipment;
- d) over-the-counter medication;
- e) prescription charges;
- f) lost income.

Productivity losses

- a) time off work;
- b) time off leisure activities;
- c) time spent on informal care by friend or family .

c. Defining missing resource use categories

Missing data for each resource use category will be explored per time point. Most questions in the resource use questionnaires have got indicator variables, for a patient to indicate whether they used a resource within the time period (YES/NO). If “Yes”, patients are asked to provide further details to allow for costing of the resource. If the indicator variable is missing or “No” but patients complete with plausible values the follow-up detail question, we will assume the indicator variable is “Yes”. If the patient indicates “Yes” but does not complete the detail follow-up question, data will be assumed missing. If both indicator variable and follow-up details are missing, the data will be considered missing, unless it is non-NHS resource use category with typically very poor completion rates, such as travel costs. In that case, we will take clinical advice and make a collegial decision on whether the data should be considered missing, zero, or plausible value assumed.

Patterns of missing data per patient over time will be assessed to ascertain whether the assumption of missing data or zero values is plausible, and assumptions will be re-evaluated otherwise.

d. Valuation of resources

Resource use collected to deliver the physiotherapy class will allow for a micro-costing approach to be taken.¹ We will take a macro-costing approach for costing delivery of treatment in both arms otherwise.

Resources in relation to the delivery of the intervention (such as staff grade and time, and resource use required to deliver care in both groups) will be valued using national UK tariffs (NHS reference costs², PSSRU³, BNF⁴) where possible. Trust finance departments will be contacted for unit costs when national tariffs not available (e.g. cost of equipment used by patients). Productivity losses will be valued using a human capital approach and ONS averaged gross weekly wages per age group and gender. When national tariffs and finance procurement costs are unavailable, we will search for other available sources, such as comparable unit costs available in the literature, or online. Unit costs and their sources will be reported. Expenses incurred privately will be valued using generic unit costs if available, or self-reported expenditure otherwise.

e. Estimation of costs and QALYs

The cost associated with each resource use item will be calculated by multiplying the units of resource used in the 12-month period by its unit cost.

Cost categories will be added by perspective: NHS, PSS, Private expenses and Productivity losses. These will be further grouped into NHS+PSS perspective for the primary analysis results and “Societal” perspective in a secondary analysis. When adding all cost categories for the societal perspective, we will deduct privately paid medications expenditure, and will use net wage rates to compute productivity losses and avoid double counting. The trial

¹ Dakin H, Abangma G, Wordsworth S. What is the value of collecting detailed costing data in clinical trials? *Trials*, 2011, 12(Suppl 1) A42.

² Department of Health (2014) NHS reference costs 2013-2014, <https://www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014> [accessed November 2015]

³ Curtis L, Unit Costs of Health & Social Care 2014, Personal and Social Services Research Unit, 2014, <http://www.pssru.ac.uk/project-pages/unit-costs/2014/> [accessed November 2015]

⁴ British National Formulary, British Medical Association and the Royal Pharmaceutical Society of Great Britain, London, [accessed November 2015]

research staff has focused their efforts into collecting NHS+PSS resource use. Our resource use questionnaire was not validated and there are generally noted difficulties for patients to understand and complete the nature of questions to collect private income and productivity losses. If private expenses and productivity losses are poorly completed, we may refrain from reporting these, or proceed with the analyses for the societal perspective results.

Quality adjusted life-years (QALYs) accumulated will be calculated from the UK preference based utility scores using the area-under-the-curve approach: assuming a linear change between the time points (2 weeks and 3, 6 and 12 months). See section II for derivation of preference based utility scores from EQ-5D-5L responses.

Cost and QALY estimates will be jointly determined in regression analysis, further adjusting by stratification variables, and baseline scores for QALYs. If feasible, and if the data distribution is non-normal, we will look at methodological ways to normalise the cost data.

f. Dealing with missing data

Economic evaluations that collect data using resource use questionnaires are prone to missing data. Cost is a compound variable adding all resource use units for each resource use category, recorded per time point. In this economic evaluation we will jointly impute cost categories and outcomes (QALYs and the primary clinical outcome: LEFS score).

A QALY will be computed based on the 5 domains of the EQ-5D questionnaire administered over 4 time points. This generates 20 variables, with potential missing data requiring imputation. Using multiple chained equation methods to impute 20 EQ-5D variables along with multiple cost categories and other outcomes of interest (LEFS), may not be computationally feasible.

We will investigate the feasibility of imputing EQ-5D domains within a multiple imputation model. Patterns of missing data in the domains of the EQ-5D will be examined. If the majority of patients only miss 1 domain, we will attempt to impute all 20 missing domains within the imputation model as this is likely to lead to more accurate imputation.⁵ If not computationally feasible, or if a majority of patients miss more than one domain, we will impute the whole utility score per time point.

⁵ Simons C L, Rivero-Arias O, Yu L M, Simon J, Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index?, *Quality of Life Research*, 2015, 24(4) 805.

We will assess the plausibility of the assumption that cost data is missing at random – assuming that missing data may not be missing completely at random, but may depend on observed variables. Missing cost categories will be imputed, jointly with the EQ-5D and the LEFs scores. Other outcome measures will be excluded to ensure the imputation model is computationally feasible to run. We will follow the statistical analysis methodology for the imputation model. We will use multiple imputation methods, using chained equations, with a minimum of 20 sets, and predictive mean matching. Costs and outcomes are compounded over time and most cost categories are more efficiently estimated summed over the 12 months period, rather than at time points. It may, therefore, not be feasible to keep the multi-level structure within the multiple imputation model. Predictive mean matching will ensure imputed missing values exist in the observed data, avoiding implausible cost and utility values. The imputation models will be stratified by trial group (treatment or standard care), treatment centre, and controlled by minimisation variables and observed patient and surgical characteristics.

For the main analysis, we will borrow complete LEFs scores from the statistical analysis to impute values for the EQ-5D and cost categories. In methodological sensitivity analysis, we will explore the effect of also imputing missing clinical outcome scores together with cost and utility values. Average cost and outcomes will be estimated using Rubin's rules and bootstrapping. We will follow the most up to date guidelines on multiple imputation methodology will be followed.⁶

g. Analysis

The primary economic analysis will be a Cost-Utility Analysis (CUA) from an NHS and PSS perspective, and will include all costs incurred at 12-months post-operative. We will separately report private expenses and loss productivity. If feasible, a secondary analysis will be conducted from the societal perspective (i.e. including all costs to patient and society).

Cost-Effectiveness Analyses (CEA) will also be carried out as secondary analyses, yielding cost per detected change in the LEFS score, from an NHS+PSS perspective, and use imputed values and intention to treat analysis in line with the primary economic analysis.

⁶ Faria R, Gomes M, Epstein D, White IR, A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials, *PharmacoEconomics*, 2014, 32(12) 193.

Costs will be adjusted for the stratification variables used in randomisation. QALYs will be adjusted for stratification variables and utility at 2-weeks post-operation in line with published guidance.⁷ The distributions for costs and outcomes will be investigated decide on methods for the independent regression analysis. We will use Seemingly Unrelated Regressions (SUR) to jointly estimate costs and outcomes. SUR methodology will also compute the correlation of residuals between the two models, and test if the two are independent or related. In other words, it will test if costs and QALYs are associated with higher or lower cost profiles for individuals, after adjusting for covariates. We will bootstrap costs and QALYs regression estimates with a minimum of 1,000 replications to account for uncertainty around the results and produce bootstrapped confidence intervals.

Costs and QALYs are already measured and compounded over a period of time. The multi-level mixed model approach for investigation of repeated measures is therefore not required for the economic result. However, should the primary statistical model use a multi-level mixed model approach to investigate the effects of surgeons or centres, the economic evaluation will also adapt to follow the statistical methodological approach.

It is possible that not all costs and benefits accruing from the intervention are observed at 12 months, if the intervention effects prolong to a longer term. We will investigate the difference in outcomes (LEFs score and QALYs) and costs from 2-weeks to 12-months post-operation in the two groups. While beyond the scope of the present analysis, we will assess whether it would be valuable to, in future research, modelling the extrapolation of the data after 12-months.

The cost-effectiveness results of this trial may be uncertain, due to structural or methodological uncertainty or sample size limitations. That is more so should the trial recruit fewer patients than expected. The economic analyses will have collected enough information to perform value of information analysis in a subsequent study to inform the decision to carry out further research, or provide evidence for decision-making based on the current trial results.

⁷ Manca A, Hawkins N, Sculpher MJ, Estimating mean QALYs in trial-based cost-effectiveness analysis: the important of controlling for baseline utility, *Health Economics*, 2005 14(5) 487.

h. Sensitivity analysis

A sensitivity analysis will be conducted to assess methodological, parameter and structural uncertainty in the economic evaluation. In particular, the following will be assessed:

1. The uncertainty around costing assumptions. These may include different sources of unit costs and, and varying local estimates. In particular, we will vary the assumptions around the costing of the intervention. We will define worse and best case scenarios, such as full capacity or low attendance in classes, and delivery of classes in a community setting. Further class size considerations are described in section IIIv.
2. The uncertainty surrounding missing data and assumptions around dealing with implausible resource use values.
3. The uncertainty around the impact of the multiple imputation model. If feasible, we will attempt a model structure where all clinical outcomes are also imputed, EQ-5D domains are imputed, and resource use by time point is imputed. These models may not be feasible.
4. We will consider the proportion of patients in standard care and the intervention, requiring additional physiotherapy and consider if of value to vary the proportion of additional physio required.

These analyses will be conducted as one-way or scenario analyses, but if data allow, and if of relevance, we may extend this to a probabilistic analysis.

i. Presentation of results

A table will be presented with the average cost of delivery of the intervention and all other mean cost categories, per trial group (treatment or standard care), per centre, at 12 months postoperative. These will include all raw available data and imputed and adjusted cost categories.

The primary economic result will be the incremental net monetary benefit (INMB) statistic, from an NHS and PSS perspective, with a range of values for the societal willingness to pay for a QALY (i.e. a year of life in perfect health). The INMB shows the incremental benefit, expressed in monetary units, less the incremental costs of our intervention over the standard care group. This requires a defined WTP threshold for the conversion of benefits (QALYs)

into monetary units. We will consider the societal willingness to pay (WTP) thresholds recommended by the National Institute for Health and Care Excellence (NICE) of £20,000 and £30,000 per QALY, as well as a published estimate of the observed willingness-to-pay (WTP) in the NHS, such as that by Claxton *et al.*⁸

We will also compute the cost-effectiveness acceptability curve (CEAC). The CEAC is a method of illustrating uncertainty surrounding cost-effectiveness results by indicating the probability that our group physiotherapy intervention is cost-effective over usual care, for a range of societal WTP values.

We will present bootstrapped costs and QALY estimates in a cost-effectiveness plane. If the intervention may be dominant (more effective and less costly than standard care) or dominated (less effective and more costly than standard care), the probability of dominance (proportion of density falling within the relevant quadrant of the CE plane) will be reported.

V. Other analyses

a. Serious adverse events

The serious adverse events (SAEs) will be presented by trial group. We will first present the total number of SAEs by trial group. We will then present the number (n=) of patients with at a SAE (irrespective of number of events experienced) by trial group with the percentage (from the number of participants randomised in each group). These percentages will be compared with Chi-square test or Fisher-exact test. Similarly, we will present the number and percentage of participants with 1 SAEs, 2 SAEs, 3 or more SAEs by trial group and we will compare those percentages with Chi-square test or Fisher-exact test.

b. Evaluation of the intervention

Following the delivery of the intervention, participants will be invited to provide their feedback on the intervention and its components.

For all continuous and score variables we will check distributions using histograms and normal plots to examine how close to normality these are before deciding which summary statistics to present. Continuous variables with approximately normal distributions will be presented as means and standard deviations (SD). Continuous variables that we anticipate

⁸ Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith PC, Sculpher M, Methods for the Estimation of the NICE Cost Effectiveness Threshold, *Health Technology Assessment*, 2013, 19 (14).

will not have an approximate normal distribution will be presented as medians and interquartile ranges (IQR). Binary/categorical variables will be presented as number (n=) and percentage (%). No imputation will be performed. Missing data will be presented as a separate category for categorical variable and the number of participants with missing information will be presented for each continuous variable.

Free text fields will be investigated with thematic analysis. Text data will be imported into Excel and then collated and coded into themes by a member of the research team. These themes will then be reviewed by a second member of the research team, and any discrepancies resolved through discussion.

c. Further exploratory investigation

Class size and the investigation of composition of intervention class effect

We will explore the impact of class composition on LEFS at 12 months by stratifying patients assigned to the intervention group into new categories defined by the number of participants in the intervention classes they attended. As this is an exploratory analysis, the classification will depend on what we observe: for example, a participant could be classified as having “only attended classes with more than 5 patients”; “only attended classes with less than 5 patients”; or “having attended some classes with less than 5 patients”. Those assigned to the usual care group will be considered as one single group as done in all the other analyses presented in this document.

Class size is an important variable to consider in the cost effectiveness analyses. The main cost categories comprising the delivery of the intervention: booked physiotherapists time and room space, are fixed costs: costs incurred irrespective of class size. If classes are not run to capacity, the delivery of the intervention would have a higher cost per patient. The primary economic analysis result will use the cost of delivery of intervention per patient attending. In sensitivity analysis (scenario or probabilistic) we will report the cost-effectiveness results using a range of class sizes, including at full capacity.

Sample characteristics and outcomes by participant subgroups

For descriptive purposes, the baseline characteristics and LEFS at 12 months will be described and compared by place of recruitment (Southmead vs Emersons Green). They will

also be reported by place of intervention delivery (Southmead <11/07/2016 vs Cosham \geq 11/07/2016) for the group of participants randomised to the intervention class. Finally, the primary outcome will be compared by method of data collection (postal questionnaire vs telephone). The trial is not powered for such analyses which should be considered as an exploratory exercise.